Geometric Models for Signal Acquisition and Processing

Chinmay Hegde

May 11, 2012

Joint work with:
Richard Baraniuk, Volkan Cevher, Marco Duarte, Kevin Kelly, Aswin Sankaranarayanan
The Data Deluge

• >250 billion gigabytes generated in 2007

Current status: digital bits > stars in the universe
> Avogadro’s number \( (6.02 \times 10^{23}) \) in 10 years
Handling Big Data

• Approach #1: Throw more resources at it
Handling Big Data

• Approach #1: Throw more resources at it
Handling Big Data (Contd.)

- Approach #2: **Model the data** in a smart manner
  - Exploit the intrinsic **physics** of the setting
  - Leverage the model to guide system design
Models are Key

$x$ → sample $N$ → compress $K$ → transmit/store

$K$ → decompress $N$ → $\hat{x}$

sparse wavelet transform

$\hat{x}$
Models are Key

\[ x \rightarrow \text{sample} \rightarrow \text{compress} \rightarrow \text{transmit/store} \rightarrow \text{receive} \rightarrow \text{decompress} \rightarrow \hat{x} \]

- Sample: \( N \gg K \)
- Compress: \( K \)
- Transmit/Store
- Receive: \( K \)
- Decompress: \( N \)
- \( \hat{x} \)

Sparse wavelet transform
Focus of this talk: Geometry of Signal Models

*Geometric intuition* enables novel methods for:

- Signal *acquisition*
- Signal *recovery*
- Signal *analysis*
Model-Based Compressive Sensing
Signal Processing Pipeline

- Established paradigm for digital data acquisition
  - sample
  - compress
  - transmit
  - reconstruct

\( \mathbf{x} \rightarrow \text{sample} \rightarrow \text{compress} \rightarrow \text{transmit/store} \)

\( \mathbf{x} \rightarrow \text{receive} \rightarrow \text{decompress} \rightarrow \mathbf{\hat{x}} \)

sparse wavelet transform
Compressive Sensing

- **New** paradigm for digital data acquisition
  - *sample and compress*
  - *transmit*
  - *reconstruct*
Compressive Sensing (CS)

\[ M \times 1 \quad \text{Samples} \]
\[ y \quad \Phi \quad x \]
\[ M \times N \]
\[ N \times 1 \quad \text{sparse signal} \]
\[ K \quad \text{nonzero entries} \]
Compressive Sensing (CS)

\[ M \times 1 \quad \text{Samples} \quad \Phi \quad N \times 1 \quad \text{sparse signal} \]

\[ M \times N \quad K \quad \text{nonzero entries} \]
Compressive Sensing (CS)

- Sampling bound
  \[ M = O(K + \log \left( \frac{N}{K} \right)) = O(K \log(N/K)) \]

- Recovery Methods
  - $\ell_1$-optimization, greedy algorithms
Signal Structure

- Sparsity: simplistic, *first-order* assumption
- Many classes of real-world data exhibit *rich, secondary structure*

wavelets: natural images

Gabor atoms: chirps/tones

pixels: background subtracted images
• **Sparse** signal:

  
  
  \[
  \begin{array}{cccccccc}
  \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} \\
  \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} \\
  \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} & \text{●} \\
  \end{array}
  \]

  \[
  x
  \]

  – only $K$ out of $N$ coefficients nonzero
Geometric Intuition

- **Sparse** signal:
  
  \[ \mathbf{x} \]
  
  - only \( K \) out of \( N \) coordinates nonzero

- **Geometry**: union of \( \binom{N}{K} \) \( K \)-dimensional subspaces aligned w/ coordinate axes

- \( N = 3, K = 1 \)
Geometric Intuition

- **Sparse** signal:
  
  - only $K$ out of $N$ coordinates nonzero

- **Geometry**: union of $\binom{N}{K}$ $K$-dimensional subspaces aligned w/ coordinate axes

- $N = 3$, $K = 2$
Sparse Signals

- Defn: \textbf{\textit{K-sparse signals}} comprise \textit{all} \(K\)-dimensional canonical subspaces.
Def: A **K-sparse union-of-subspaces model** comprises a particular (*reduced*) set of $L_K$ $K$-dim canonical subspaces.
Sampling Bounds

- **RIP:** stable embedding

\[ M = O(K \log(N/K)) \]

[CRT06, D06, BDDW08]
Sampling Bounds

- **Model-RIP:** stable embedding

\[ M = O(K + \log(L_K)) \]

[BD09, BCDH10]
Iterated Thresholding

- goal: given $y = \Phi x$, recover $x \in \Sigma_K$

initialize $i = 0, \ x_0 = 0$

iterate:
  - $\hat{x}_{i+1} \leftarrow \text{thresh}(\hat{x}_i + \Phi^T(y - \Phi x_i))$

return $\hat{x} \leftarrow \hat{x}_i$

[BD08]

K-planes
Iterated Model Thresholding

• goal: given $y = \Phi x$, recover $x \in \mathcal{M}_K$

initialize $i = 0$, $x_0 = 0$

iterate:
  • $\hat{x}_{i+1} \leftarrow \mathcal{M}(\hat{x}_i + \Phi^T(y - \Phi \hat{x}_i))$

return $\hat{x} \leftarrow \hat{x}_i$

[BCDH10]
Recovery Guarantee

Suppose we observe

\[ y = \Phi x^* + e, \quad x^* \in \mathcal{M} \]

Then, the estimates of Iterated Model Thresholding satisfy:

\[ \| \hat{x}_i - x^* \|_2 \leq 2^{-i} \| x^* \|_2 + 15 \| e \|_2 \]
Wavelet Sparsity

- Typical of wavelet transforms of natural signals and images (piecewise smooth)
**Tree-Sparsity**

- **Model:** $K$-sparse coefficients + significant coefficients lie on a **rooted subtree**

- Typical of wavelet transforms of natural signals and images (piecewise smooth)
Tree-Sparsity

- **Model**: $K$-sparse coefficients
  - significant coefficients lie on a rooted subtree

- **Tree-sparse approx**: find best rooted subtree of coefficients
  - CSSA [B] $O(N \log N)$
  - dynamic programming [Donoho] $O(N)$
Tree-Sparsity

- **Model:** $K$-sparse coefficients
  - significant coefficients lie on a rooted subtree
Other Structured Sparsity Models

- Block-sparsity
- $\Delta$-separated spikes
- Markov Random Fields

[BCDH10, HDC09, CDHB08, CIHB09, HB11]
Manifold Models

- $K$-dimensional parameter vector captures degrees of freedom in signal $x \in \mathbb{R}^N$

$$x = x(z), \quad z \in \mathbb{R}^K$$
Sampling Bounds

\[ M = \mathcal{O} \left( \frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2} \right) \]
Recovery

\[ \hat{x}_{i+1} \leftarrow \mathcal{M}(\hat{x}_i + \Phi^T(y - \Phi\hat{x}_i)) \]

[SC11]
Manifold-Based Recovery

- Real-data experiments with the Single-Pixel Camera

Joint work with Kelly Lab
Signal Separation and Denoising
Signal Separation

- Cocktail party problem
Signal Separation

- Cocktail party problem
- Audio click removal
Signal Separation

- Cocktail party problem
- Audio click removal
- Morphological components analysis (MCA)
Signal Separation

- Cocktail party problem
- Audio click removal
- Morphological components analysis (MCA)
- ...

- Numerous settings have been explored
  - spike and sines
  - incoherent bases
  - “robust” recovery in compressive sensing
  - low-rank + sparse matrix decomposition
Model

• Signal of interest:

\[ x = a^* + b^*, \quad a^* \in A, \ b^* \in B \]

• Noisy linear observations:

\[ y = \Phi x + e = \Phi(a^* + b^*) + e \]
• Key concept: **incoherence** (b/w manifold secants)

\[
\left| \left\langle \frac{a_1 - a_2}{\|a_1 - a_2\|}, \frac{b_1 - b_2}{\|b_1 - b_2\|} \right\rangle \right| \leq \epsilon
\]
Successive Projections onto Incoherent Manifolds (SPIN)

• goal: given \( y = \Phi(a^* + b^*) + e \), recover \( (a^*, b^*) \)

initialize \( i = 0, \ x_0 = 0 \)

iterate:

• \( a_{k+1} \leftarrow \mathcal{P}_A(a_k + \eta \Phi^T(y - \Phi(a_k + b_k))) \)
• \( b_{k+1} \leftarrow \mathcal{P}_B(b_k + \eta \Phi^T(y - \Phi(a_k + b_k))) \)

until convergence

[HB12]
Multi-Manifold Recovery

- $N = 64 \times 64 = 4096$, $M = 50$

- Near-perfect recovery with $M/N = 1.2\%$ meas.!
Matrix Decomposition
Matrix Decomposition

- Low-rank + Sparse model

\[ X = L + R + S \]
Matrix Decomposition

- Low-rank + Sparse model

\[
X = L + S
\]

- Efficient projection operators exist for both manifolds

- **Violates** global incoherence assumption
Numerical Results - SPIN

\[ \rho = \frac{K}{M} \]
\[ \delta = \frac{M}{N^2} \]
Linear Dimensionality Reduction
"Manifold Learning" - Isomap, LLE, MVU, ...
- Obtain a low-dimensional representation of the data
- Preserve geometric information
Dimensionality Reduction

• Despite their great promise, nonlinear methods suffer from drawbacks:
  – (often) do not generalize to out-of-sample points
  – (often) unstable

• Alternative: linear dim. reduction methods

\[ X \mapsto AX \]

– Principal components analysis (PCA), and variants
– Random projections
Linear Dim. Reduction Methods

• Principal Components Analysis (PCA)
  – Easy to compute
    \[ X = U S V^T \]
  – But **distorts** pairwise distances

• Random Projections
  – Guarantees pairwise distance preservation (JL)
    \[ 1 - \delta \leq \frac{\|\Phi(x_1 - x_2)\|^2}{\|x_1 - x_2\|^2} \leq 1 + \delta \]
  – But constants are **poor**
  – Oblivious to structure of data
“Good” Linear Maps

$X \subset \mathbb{R}^N$

$\mathcal{M}$
“Good” Linear Maps

- Goal: preserve norms of all pairwise secants of $X$
- If $X$ is a dense enough sampling, then $\Phi$ is an isometric mapping of $\mathcal{M}$

\[ 1 - \delta \leq \frac{\|\Phi(x_1 - x_2)\|^2}{\|x_1 - x_2\|^2} \leq 1 + \delta \]
Designing a “Good” Linear Map

Want: a short, fat matrix $\Phi$, such that

$$1 - \delta \leq \|\Phi v_i\|^2 \leq 1 + \delta$$

$i = 1, 2, \ldots, Q$

$\updownarrow$

minimize $\text{rank}(\Phi)$, subject to

$$\left|\|\Phi v_i\|_2^2 - 1\right| \leq \delta$$

$i = 1, 2, \ldots, Q$
The Lifting Trick

- Convert quadratic constraints in $\Phi$ into linear constraints in $P = \Phi^T \Phi$

- Use a nuclear-norm relaxation of the rank

- **Simplified** problem:

  $$\begin{align*}
  \text{minimize} & \quad \|P\|_* \\
  \text{subject to} & \quad \|A(P) - 1\|_\infty \leq \delta \\
  & \quad P \succeq 0, \quad P = P^T
  \end{align*}$$
A Fast Algorithm

- Alternating Direction Method of Multipliers (ADMM)
  - Introduce auxiliary variables
    \[
    \text{minimize } \|P\|_\star \\
    P = L, \ A(L) = q, \ \|q - 1\|_\infty \leq \delta
    \]
  - Every iteration decouples into three optimizations, each evaluated in \textbf{closed-form}
  - Geometric intuition: successive \textbf{projection} onto suitably defined convex sets
MNIST Dataset
M=20 linear measurements enough to ensure isometry constant of 0.01!
Classification: Circles & Squares

[Image of circles and squares]

[Graph showing probability of error vs. number of measurements M]

[Reference: HSB12]
Summary

• Models are the key

• One (nice) way to study models: **Geometry**

• Geometric intuition enables potential novel methods for signal
  – acquisition (linear dim. reduction)
  – reconstruction (model-CS)
  – analysis (source separation)

• Advantages of the geometric approach:
  – Embrace nonlinearity, non-convexity
  – Concise framework for characterizing limits of systems
  – Unifies, generalizes
What’s Next?

• Re-imagining the sig. proc. pipeline: **Adaptivity**
  – Design measurements according to signal
  – Closed loop sensing + reconstruction

• Pushing the limits
  – Do things work when almost all the data is bad
  – What kind of questions to ask?

• Beyond signals and images
  – rankings, “likes”, questionnaires, networks, etc.
References


